Assessment and prediction of the impact of road transport on ambient concentrations of particulate matter PM10

Suleiman, A.; Tight, M. R.; Quinn, Andrew

DOI:
10.1016/j.trd.2016.10.010

Assessment and prediction of the impact of road transport on ambient concentrations of particulate matter PM_{10}

A. Suleiman*

alaminsjam@hotmail.com

axs526@bham.ac.uk

M.R. Tight

A.D. Quinn

School of Engineering, University of Birmingham, Edgbaston, Birmingham B15 2TT, United Kingdom

*Corresponding author.

Abstract

The main challenge facing the air quality management authorities in most cities is meeting the air quality limits and objectives in areas where road traffic is high. The difficulty and uncertainties associated with the estimation and prediction of the road traffic contribution to the overall air quality levels is the major contributing factor. In this paper, particulate matter (PM_{10}) data from 10 monitoring sites in London was investigated with a view to estimating and developing Artificial Neural Network models (ANN) for predicting the impact of the road traffic on the levels of PM_{10} concentration in London. Twin studies in conjunction with bivariate polar plots were used to identify and estimate the contribution of road traffic and other sources of PM_{10} at the monitoring sites. The road traffic was found to have contributed between 24% and 62% of the hourly average roadside PM_{10} concentrations. The ANN models performed well in predicting the road contributions with their R-values ranging between 0.6 and 0.9, FAC2 between 0.6 and 0.95, and the normalised mean bias between 0.01 and 0.11. The hourly emission rates of the vehicles were found to be the most contributing input variables to the outputs of the ANN models followed by background PM_{10}, gaseous pollutants and meteorological variables respectively.

Keywords: Road traffic contribution; Bivariate polar plot; Artificial neural network; Particulate matter

1 Introduction

Emissions from traffic make tremendous contributions to overall pollutant concentrations in urban areas. The Particulate matter, especially with an aerodynamic diameter less than ten (PM_{10}), is one of the pollutants that is majorly contributed to by road traffic. These small particles have been shown to be detrimental to human health and the urban environment (Anderson et al., 2012; Lawal et al., 2015; Brunekreef et al., 2009). Particulate matter also contributes to visibility impairment (Yang et al., 2012) and often cause road accidents (Abdel-Aty et al., 2011). The EU directive on ambient air quality and cleaner air for Europe (2008/50/EC), puts a limit to the level PM_{10} that should not be exceeded for the purpose of reducing its impact on human health. The EU Directives state that the daily and annual mean of PM_{10} should not exceed 50 $\mu g/m^3$ and 40 $\mu g/m^3$ respectively. The daily limit should not be exceeded more than 35 days a year. These limits can be effectively controlled if the sources of particulate matter and the factors affecting its levels are adequately characterised and quantified.

The levels of particulate matter in some urban areas in the UK and other European countries are declining due to improvements in the technology of vehicles and the implementation of the air quality objectives and directives. However, emission in areas close to major roads remained the challenge of the regulatory authorities due to frequent cases of exceedances of air quality limits and objectives. Particulate emission from roads can both emanate from the vehicle exhaust or non-exhaust sources such as wear and tear of vehicle parts (e.g. tyre and clutch). Other important sources of road traffic particulate emission are re-suspension of dust due to vehicles movement (Pant and Harrison, 2013) and brake pads. Hence the need for more studies on the accurate characterisation, estimation and prediction of road traffic contribution to particulate concentrations in the urban area. The information obtained from such studies could be useful in identifying relevant and effective control measures to the dominant sources of the particles in a particular area e.g. proximity to major roads.

Several studies have been carried out on the identification of sources of ambient pollutants using Principal Component Analysis (PCA), Positive Matrix Factorization (PMF) and conditional Bivariate probability function (Singh et al., 2013; Pokorná et al., 2015; Uria-Tellaetxe and Carslaw, 2014). Pant and Harrison (2013) reviewed and discussed the pros and cons of the methodologies for the assessment of road traffic emissions and the receptor modelling of particulate matter. The assessment methods include direct measurements near roads/highways and tunnels, twin studies, dynamometer tests and tracer studies (Gouriou et al., 2004; Jones and Harrison, 2006; Chiang et al., 2012; Ketzel et al., 2003). The receptor modelling methods discussed by Pant and Harrison (2013) include multivariate statistical methods (i.e. PCA, PMF, UNMIX, and Multilinear Engine (ME)), and Chemical Mass Balance (CMB) model. The multivariate statistics and CMB methods are quite revealing and informative. However, they require the use of expensive instruments in measuring relevant tracer elements associated with the different sources of the particles. These expenses make them difficult to be applicable for a city-wide study. Moreover, the twin studies are easy to apply, even though Harrison et al. (2004) discovered that the site geometry and the pattern of the circulation flow at the sites affect the estimates of the ambient concentrations of particulate matter. These effects can be effectively managed when the twin studies are applied in conjunction with bivariate polar plots to reduce the uncertainty of the estimates due to the nature of air flows and geometry of the sites (Masiol and Harrison, 2015; Carslaw and Beevers, 2013; Carslaw et al., 2006). The main aim of this paper is to assess and quantify the impacts of some major roads on the ambient concentrations of PM_{10} in London. Moreover, to develop artificial neural network based models for predicting the contribution of the roads on the roadside particulate matter concentrations.

2 Materials and ~~M~~methods

2.1 Monitoring sites

The monitoring sites selected for this study are located in London and are mostly maintained by the London boroughs, Department for Environment, Food and Rural Affairs (DEFRA), and Transport for London (TfL). The location of the sites is shown in Fig. 1. The sites are categorised into a kerb, roadside, urban background, and rural background sites. The roadside and kerb sites are located between 1 ~~to~~and 10 m from the major roads. HK6, IS2, KC5, and MY1sites are situated in the street canyons. GR5, GR8, KC2, CR4 and CD3 sites are located at junctions while BT4 and GR8 sites are situated in an open area (see Table 1).

Fig. 1 Googlestreet map showing the locations of MY1 and BL0 sites.

Table 1 Properties of the London monitoring sites.

Site code	Easting	Northing	Site name	Site type	Distance to the road (m)	Traffic volume (veh/h)	Average PM_{10} ($\mu g/m^3$)	PM_{10} (%) available
BL0	530,123	182,014	Camden - Bloomsbury	Urban background			21.76	92.1
BT4	520,866	185,169	Brent - Ikea	Roadside	Not available	4389	43.25	90.4
CD3	530,057	181,285	Camden - Shaftesbury Avenue	Roadside	3	1700	34.00	91.4
CR3	532,336	168,934	Croydon - Thornton Heath	Suburban			21.13	91.0
CR4	532,583	165,636	Croydon - George Street	Roadside	8	2500	25.00	95.0
CT3	533,480	181,186	City of London - Sir John Cass School	Background			27.51	91.5
GR4	543,978	174,655	Greenwich - Eltham	Suburban			21.91	99.6

GR5	538,960	177,954	Greenwich - Trafalgar Road	Roadside	5	1500	23.37	99.6
GR8	540,200	178,367	Greenwich - Woolwich Flyover	Roadside	3	7000	40.00	97.3
HK6	532,947	182,575	Hackney - Old Street	Roadside	6	2500	31.83	94.1
IS2	530,698	185,735	Islington - Holloway Road	Roadside	3	2000	30.73	98.6
IS6	531,325	186,032	Islington - Arsenal	Urban background			22.40	97.5
KC1	524,046	181,750	Kensington and Chelsea - North Ken	Urban background			21.11	96.7
KC2	526,527	179,646	Kensington and Chelsea - Cromwell Road	Roadside	4	2800	33.71	81.4
KC5	525,671	179,080	Kensington and Chelsea-Earls Court Rd	Kerbside	Not available	1600	35.83	98.9
MY1	528,125	182,016	Westminster - Marylebone Road	Kerbside	1.5	3327	43.25	97.5

The hourly average traffic volume on these roads ranges between 1500 veh/h and 7955 veh/h. The BL0, CR3, CT3, GR4, HA1, IS6 and KC1sites are either urban or rural background monitoring sites. The background sites are mostly located in areas where there is less influence of local pollution sources. The data collected at the sites include particulate matter (PM_{10}, $PM_{2.5}$) and gaseous pollutants (NOx, NO_2, NO, SO_2, CO, and O_3), traffic volume and speeds. Others are meteorological variables (wind speeds, wind direction in-street flows, solar radiation, relative humidity and ambient temperature). The instruments used for the monitoring of $PM_{2.5}$ and PM_{10} at most of the sites in London include two similar Tapered Element Oscillating Microbalances (TEOM) Model 1400AB with different sampling heads design (Aurelie and Harrison, 2005). The filter dynamics measurement system (FDMS) have also been installed at some of the stations to minimise the problems of loss associated with the TEOMs. Some small number of sites are using β-attenuation analysers for the measurements. Data from these monitoring sites are being measured according to EU protocols and are undergoing the quality assurance and quality controls according to Automatic Urban and Rural Network (AURN) and London Air Quality Network (LAQN) standards. The data is openly accessible through the London Air Archives (LondonAir, 2013) and UK Air Quality Archive (UK-AIR, 2013). The meteorological data in London was collected from London Heathrow Airport Meteorological station through BADC data services (MIDAS Land Surface, 2013).

2.2 Data

The hourly average mean PM_{10} concentrations at the sites for a period between 2000 and 2012 were between 22.7 μg/m^3 at GR5 and 43.33 μg/m^3 at MY1. In most of the sites, the hourly average PM_{10} concentrations were under the EU annual mean limit value of 40 μg/m^3. The 90.4th percentile of the PM_{10} concentrations at the sites ranged from 44.1 μg/m^3 at GR5 to 72.4 μg/m^3 at GR8. The percentages of missing data in all the sites selected were less than 10% except at KC2 where the missing data was up to 19%. The hourly average wind speed measured at Heathrow airport between 2000 and 2012 was 2.1 m/s and the 95th percentiles, and the maximum wind speeds were 9.9 m/s and 4.3 m/s respectively. The prevailing winds were from the Southwest and West directions. The directions of the dominant winds at the sites govern the location of the air quality monitoring sites. For example, at Marylebone Road, the air quality monitoring site is located to the south of the road to take the advantage of the effect of across Canyon vortex usually caused by the prevailing wind. The prevailing wind makes the flow circulate within the street canyon and deliver most of the pollutants to the leeward side of the street canyon (Tomlin et al., 2009).

2.3 Quantification of upper limit of road traffic contribution

The method adopted for the estimation of the upper limit of road traffic contribution follows the method developed by Carslaw et al. (2006) to quantify the contribution of aircraft and other on airport sources to ambient oxides of nitrogen. The same method was also applied by Masiol and Harrison (2015) to estimate the impact of the Heathrow Airport and the M25 and M4 motorways on the surrounding air quality. The method involves estimation of road PM_{10} concentration increment by subtracting the background concentration upwind of the road site (i.e. twin studies) and the use of bivariate polar plots to locate the wind sectors related to the source in question. The contribution of a source will then be estimated by filtering the data by the time of the activities at the source, the wind sector, and the wind speed. The Bivariate polar plots describe the joint variation of pollutant concentrations, wind speeds and wind direction on a continuous surface using polar coordinates (Carslaw and Beevers, 2013). In their separate studies, Carslaw et al. (2006) and Masiol and Harrison (2015) estimated the upper limit of airport contributions considering the appropriate wind sectors and wind speeds greater than 3 m/s to eliminate the influence of local sources such as roads. Conversely, in this study we are interested in the contributions of the local sources hence we used the data covering 6:00–22:00 associated with the wind sectors related to the roads and with wind speeds less than 3 m/s to isolate the influence of other sources far away from the monitoring units. The pollutants, traffic, and meteorological data associated with the estimates obtained were used to train Artificial Neural Network (ANN) models for the prediction of the upper limit of the contribution of the roads.

2.4 Artificial Neural Network (ANN) modelling

ANN models are designed to mimic the behaviour of the human brain. The human brain is made up of interconnected synaptic neurons that are capable of learning and storing information about their environment (Bishop, 1995). A neuron model is made up of three elements, the connecting links characterised by their strength and a linear combiner that combines the weighted input signals. Moreover, it has an activation function for limiting the amplitude range of the neuron's output to some finite value. The commonly used ANN method is Multilayer Perceptron Network (MLPN) trained using a back propagation algorithm (BP). The MLPN method involves designing an appropriate neural network architecture consisting of serially interconnected layers, training the network on a training data and testing the network on a test data set. The network layers include input layer where the input variables are received and the hidden layer where the sum of the weighted inputs from the input layer are received through the connecting links of various weights. The weighted inputs are transformed into a higher dimension using the hidden layer activation function (e.g. sigmoid function). The last layer is the output layer where the outputs of the hidden layer are received through connecting links, and the final output of the network are estimated using output layer activation functions usually linear. The network outputs are then compared with the target samples, and the errors are estimated and propagated back to update the previously estimated network weights. This procedure is repeated until the network with minimum error, and good generalisation is obtained. The main task in the design of the neural network is the determination of an appropriate number of hidden neurons and the selection of input variables that will produce a model with the desired generalisation and prediction accuracy.

The ANN methods have been successfully applied in many air quality studies (Taspinar, 2015; Ragosta et al., 2015; Elangasinghe et al., 2014a). They were used in studies involving predicting and forecasting of air pollutants ranging from the current hour to several days in advance (Russo et al., 2013; de Gennaro et al., 2013). In an attempt to tackle the challenges of training ANN models, many studies involving neural network often used cross-validation or evolutionary algorithms such as Genetic algorithm and particle swarm optimisation methods to derive an optimum architecture for the ANN models ((Ding et al., 2011a,b; He et al., 2014). In addition, the input selection methods often used in air quality studies with ANN include PCA (Taspinar, 2015; Ragosta et al., 2015), stepwise regression (Russo et al., 2013; Lima et al., 2013), and cluster analysis (Elangasinghe et al., 2014b).

In this paper, the Multilayer Perceptron (MLP) was used to train the ANN models for predicting the upper limit of road traffic contribution to the roadside particulate matter concentrations. The optimum architecture of the ANN was determined using a model tuning function (*train*) provided in the *caret* package of an R statistical software (R Development Core Team, 2015; Kuhn, 2012). The function uses the resampling method specified (e.g. bootstrap) to hold out certain sample of the training data and then fit the model on the remainder of the samples over a specified range of the model parameters (i.e. number of hidden layer neurons and weight decay). The hold-out samples would then be used to test and evaluate the model performance. The function uses RMSE values to determine the optimal model parameters and then fit the final model to all the training data-sets using the optimal parameters. The selected models were then tested using the test data set and evaluated using statistical performance metrics including coefficient of correlation R, mean bias and normalised mean bias (MB, NMB), mean gross error and its normalised form (MGE, NMGE). Others were the root mean squared error RMSE, the coefficient of efficiency COE and index of agreement IA. The models were also compared with the observation of a time variation plots to examine how they accurately captured the temporal variations in the original observations. The relative importance of the input variables to the outputs of the ANN models were estimated using connection weight approach (Olden and Jackson, 2002). The method involves calculating the products of the input – hidden and hidden – output connection weights for each input and output neurons and sum the products across all hidden neurons (Olden et al., 2004).

3 Results and discussion

3.1 Spatial analysis of the PM_{10} concentrations using Bivariate Polar Plots (BPP)

The background sites selected for each roadside site were located upwind to avoid the influence of the road sources under consideration. The sites are also located mostly in areas where there is the less likely influence of other local PM_{10} sources. For example, the background site selected for MY1 was BL0 which is located approximately 2 km to the east of MY1 in a Russel park square London (see Fig. 1).

The Bivariate polar plots of the total roadside PM_{10} concentrations at most of the sites show that both local and far distance sources have an influence on the elevated PM_{10} concentrations at the sites (see Fig. 2 first panel for MY1 site). The plots revealed that the higher concentrations are associated with both higher and lower winds from several directions. The MY1 monitoring unit is located on the southern side of Marylebone Road in a street canyon formed by high-rise buildings of different heights in the city of Westminster London. The site is situated few metres away from the junctions connecting Luxborough Street from the east and Baker Street from the west. The road which runs along the southwest to the northeast axis is one of the busiest roads in central London with the traffic volume of over 4000 veh/h and the hourly average PM_{10} concentrations of about 43.25 µg/m^3. The higher mean PM_{10} concentrations ranging from 45 to 70 µg/m^3 at the site are more related to the winds along the road and the recirculating flows within the canyon as shown in the first panel of Fig. 2.

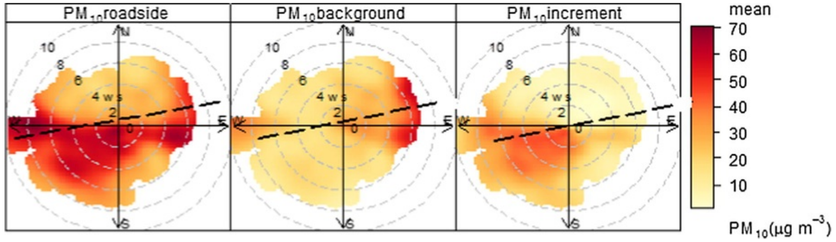

Fig. 2 Bivariate polar plots of PM_{10} concentration showing the effect of background sites on the roadside site.

The BPP of the background $PM1_0$ shown in the second panel of Fig. 2 indicates that the higher PM_{10} concentrations at BL0 site, are more related to the higher winds coming from east, indicating the prevalence of the concentrations from long range transport. The PM_{10} road increments obtained as the difference between the roadside and background PM_{10} concentrations are more related to the lower winds and the concentrations reduces with an increase in wind speed as shown in the third panel of Fig. 2. The behaviour shown by the PM_{10} road increment shown in Fig. 2 is expected from the ground level sources e.g. traffic. However, at higher wind speeds, say >3 m/s, PM_{10} from far distant sources might be transported to the site. Therefore, PM_{10} increments associated with the higher winds are isolated to obtain the upper limits of the road contribution to the total PM_{10} concentrations at the monitoring sites. Although higher wind speeds could result in higher resuspension of the particles and formation of street canyon vortices, Kassomenos et al. (2012) reported that the non-wind driven component of the re-suspended coarse particles ($PM_{2.5}$–PM_{10}) dominates with around 74–90% of the total coarse particle mass in the three examined cities.

Therefore, the BPP for PM_{10} concentrations have been drawn for each of the ten roadside monitoring sites to qualitatively access the behaviour of PM_{10} road increments at the sites. For each monitoring site. four BPPs were derived, each for the total concentrations, the roadside increment, the roadside increment associated with wind speed less than 3 m/s and the roadside increment related to wind speed greater than 3 m/s as shown in Fig. 3a, b, c and d respectively. The surface of BPP plots is determined by the combination of the wind speed, wind direction and the PM_{10} concentrations. The circles indicate the wind speeds and the colour coded surface show the spread of the PM_{10} concentrations at various wind directions. Fig. 3c shows a complete surface indicating that at lower wind speed the concentration of PM_{10} is associated with all the wind directions. However, at higher winds the PM_{10} concentrations are only associated with the directions of the prevailing wind directions as shown in Fig. 3d.

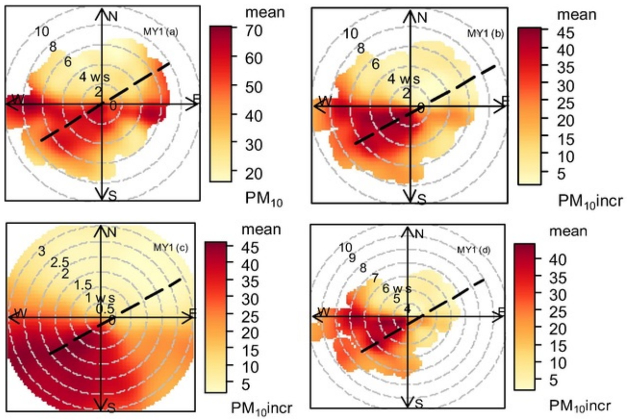

Fig. 3 Bivariate polar plots of PM_{10} concentration at the MY1 monitoring site showing the changes in concentrations due to background concentrations and wind speed.

The BPP of the total PM_{10} increments shown in Fig. 3b revealed that the higher PM_{10} increment ranging from 25 to 45 µg/m^3 are more related to the winds coming from the west and southwest indicating the prevalence of the canyon channelling flows. This behaviour is more pronounced when the increments related to the lower winds are isolated as shown in Fig. 3c. However, at higher winds, the flows along the canyon from the southwest might carry the higher concentrations from the Baker Street junction located to the west through the road down to the monitoring unit as shown in Fig. 3d.

Similarly, the BPPs of the total PM_{10} increments for BT4, CD3, and CR4 sites indicate that the increments are mostly influenced by local sources because they are more related to the lower wind speeds and decreases with an increase in wind speeds. Conversely, at the remaining sites, the higher concentrations associated with both low and high wind speeds indicating the effect of both local and far distance sources. There are two things to consider at the sites where increments are associated with both higher and lower winds. First, the background concentrations might have been affected by the influence of local sources. Therefore, it is not a true representative of a background concentration. Second, the concentrations might have been transported to the monitoring sites by the higher winds flowing along the street and by the circulation vortices in the case of street canyons. The BPPs obtained separately for higher winds, and lower winds indicate that most of the higher concentrations related to the higher wind speeds are carried to the monitoring units by the channelling flows and in some cases by the recirculation flows. However, the concentrations associated with the lower winds shows that both the channelling flows and the recirculation vortices play a vital role, and the concentrations are always associated with the direction of the location of the roads and the directions opposite to the road in a case where canyon vortices are formed. Therefore, the concentrations at lower winds that are associated with the wind direction related to the location of roads were considered the upper limit contribution of the road traffic on the roads.

3.2 Quantification of the road traffic contribution to the roadside PM_{10} concentrations

The bivariate polar plot analysis of the PM_{10} concentrations presented in Section 3.1 provided us with the qualitative assessment of the likely effect of the roads near the monitoring stations. This section went further to provide a quantitative estimate of the upper limit of the road traffic contribution using a similar approach to that employed by Carslaw et al. (2006) and Masiol and Harrison (2015) for quantifying the upper limit contribution of Heathrow airport. The method involved identifying the appropriate site pairs i.e. roadside and background sites and then estimating the roadside increment by subtracting the background concentrations from the roadside concentrations. The concentrations collected between 06:00 and 22:00 were then extracted and divided according to those associated with wind speed greater than 3 m/s and less than 3 m/s. Furthermore, the BPP of the PM_{10} increment for each wind speed class and the combined data for the period was derived. The BPPs were then used to identify the wind sectors related to the roads that contribute most to the elevated concentrations at the sites. The data for the selected wind sectors were then extracted, and their mean and frequency were obtained. The upper limit of the road contribution was estimated as the mean PM_{10} increment related to the selected wind sectors and wind speed less than 3 m/s. The period between 6:00 and 22:00 and the wind speed less than 3 m/s were chosen to maximise the road contribution and eliminate the influence of long distance sources respectively (Carslaw et al., 2006). The results of the estimates are shown in Table 2.

Table 2 Estimates of the contribution of road traffic and other sources to the concentrations of PM_{10}.

Site pairs	Wind speed (m/s)	Wind sector (degrees)	Total mean PM_{10} (µg/m³)	Total mean PM_{10}inc (µg/m³)	Mean PM_{10}inc (µg/m³)	Percent road contribution (%)	Percentage of observations (%)
BT4 – KC1	Combined	0–140	37.04	14.51	22.79	62%	9%
BT4 – KC1	>3 m/s	30–130	37.04	14.51	21.76	59%	2%
BT4 – KC1	<3 m/s	40–230	37.04	14.51	22.07	60%	27%
CD3 – KC1	Combined	75–200	34.22	11.6	12.16	36%	49%
CD3 – KC1	>3 m/s	65–200	34.22	11.6	10.52	31%	12%
CD3 – KC1	<3 m/s	90–255	34.22	11.6	14.86	43%	21%
CR4 – HA1	Combined	70–255	25.62	8.51	12.16	47%	26%
CR4 – HA1	>3 m/s	70–240	25.62	8.51	9.27	36%	5%
CR4 – HA1	<3 m/s	90–180	25.62	8.51	12.81	50%	22%
GR5 – GR4	Combined	60–90	23.37	1.42	3.69	16%	18%
GR5 – GR4	>3 m/s	60–90	23.37	1.42	3.37	14%	2%
GR5 – GR4	<3 m/s	340–90	23.37	1.42	5.5	24%	7%
GR8 – GR4	Combined	150–310	40.63	18.68	22.86	56%	27%
GR8 – GR4	>3 m/s	250–330	40.63	18.68	24.36	60%	4%
GR8 – GR4	<3 m/s	150–310	40.63	18.68	24.15	59%	35%
HK6 – CT3	Combined	250–340	31.83	4.45	10.82	34%	18%

HK6 – CT3	>3 m/s	90–250	31.83	4.45	9.21	29%	4%
HK6 – CT3	<3 m/s	90–320	31.83	4.45	12.48	39%	17%
IS2 – IS6	Combined	140–330	30.73	8.26	3.04	10%	11%
IS2 – IS6	>3 m/s	180–335	30.73	8.26	1.66	5%	2%
IS2 – IS6	<3 m/s	145–330	30.73	8.26	7.76	25%	21%
KC2 – KC1	Combined	60–300	33.71	11.96	12.52	37%	39%
KC2 – KC1	>3 m/s	60–300	33.71	11.96	15.59	46%	2%
KC2 – KC1	<3 m/s	140–280	33.71	11.96	12.71	38%	31%
KC5 – KC1	Combined	190–270	35.83	14.61	13.62	38%	7%
KC5 – KC1	>3 m/s	0–90	35.83	14.61	11.95	33%	2%
KC5 – KC1	<3 m/s	50–140	35.83	21.65	16.39	46%	44%
MY1 – BL0	Combined	140–270	43.33	21.65	26.89	62%	35%
MY1 – BL0	>3 m/s	170–270	43.33	21.65	24.34	56%	4%
MY1 – BL0	<3 m/s	150–270	43.33		25.2	58%	35%

The first column displays the site pairs; the second column displays the wind speed classes while the third column shows the selected wind sectors for each wind speed class. The overall hourly mean PM_{10} concentrations and the hourly average roadside increment at wind speed/wind sector are shown in third and fourth columns respectively. The fifth column shows the percentage of the mean increment at each wind speed/wind sector as a proportion of the average PM_{10} concentrations for the whole observation over the study period. The sixth column is the percentage of the observation for each wind speed/wind sector as a proportion of the total observations at the site.

The upper limit of the road contribution constitutes between 24% and 62% of the mean PM_{10} concentrations at the sites. The sites with the higher hourly average traffic volume seem to have higher contributions irrespective of their locations. For example, BT4, GR8, and MY1 contributed about 58–60% of the mean PM_{10} concentrations recorded at their respective locations as shown in Table 2. The frequency of observations associated with the upper limit estimation constitutes about 21–44% of the total observation at the sites. There was no much difference between the average contribution of other sources and the roads in terms of the level of concentrations. However, there is an enormous difference in the frequency of their respective observations. It was observed that the frequency of observations associated with the higher wind speeds at wind sectors related to the roads constitute only about 2–12% of the total observation as shown in Table 2. In the UK, the average road transport contribution to the overall PM_{10} emission is about 27% (DEFRA, 2013), however in London, road transport contributes about 49% of the total PM_{10} emission (TfL, 2014). The hourly average upper limit of the PM_{10} increment estimated in this study was about 15.39 µg/m^3 which is about 46% of the average PM_{10} concentrations observed at the sites. This nearly corresponds to the 49% estimated actual road transport contribution by the London Atmospheric Emissions Inventory (LAEI) (TfL, 2014). Although the results obtained in this study is comparable to LAEI estimates, the actual road traffic contribution is not known. Also, the assumptions about the wind speed and wind direction dependencies of the PM_{10} concentrations, though, shown to be robust, might affect the accuracy of the estimates since the boundaries of the wind parameters are just estimates.

3.3 ANN modelling

The input data selected for the ANN modelling comprises of the background PM_{10}, roadside NOx, NO2 and SO_2. The traffic variables included are the PM_{10} emission rates of the eight categories of traffic composition i.e. petrol cars, diesel cars, taxi, LGV, Rigid HGV, articulated HGV, Bus and Coach and motorcycle. Others are the meteorological variables including Rainfall, Relative humidity, solar radiation, Temperature, Barometric Pressure, Wind Speed and Wind directions. The input variables to the ANN models were first processed using Principal Component Analysis (PCA). Pre-processing the data in this way, allows for the derivation of uncorrelated variables (i.e. PCs) to reduce the dimensionality of the input space which will enhance the performance of the ANN models to be developed. The first PCs that explained 99% of the variance in the data were selected as the model's inputs. The first two PCs contributed about 57% of the total variation in the data (see Table 3). The first PC is dominated by the contribution of traffic variables and has highlighted the positive correlation between the traffic variables and the pollutant concentrations. The second PC show high correlation with the pollutant and meteorological variables in the data. Most importantly, the second PC displays the negative correlation between the wind speed, solar radiation and temperature on one side and the pollutant variables on the other side.

Table 3 Results of the principal component analysis of the input data.

Principal components	Eigenvalue	Percentage of variance	Cumulative percentage of variance
Comp 1	7.93	41.73	41.73
Comp 2	2.99	15.72	57.45
Comp 3	2.11	11.10	68.55
Comp 4	1.13	5.92	74.47
Comp 5	1.07	5.61	80.07
Comp 6	0.87	4.60	84.67
Comp 7	0.71	3.71	88.39
Comp 8	0.66	3.46	91.85
Comp 9	0.53	2.78	94.63
Comp 10	0.38	1.98	96.61
Comp 11	0.31	1.63	98.23
Comp 12	0.19	1.02	99.26
Comp 13	0.10	0.51	99.77
Comp 14	0.03	0.15	99.92
Comp 15	0.01	0.07	99.99

The negative correlation shows that when the temperature is low, the pollution level tends to rise, and also the pollution levels reduces with an increase in the ventilation of the urban area. Low temperatures contribute to high PM_{10} concentrations through condensation of volatile compounds and possibly because of residential heating. Moreover, low temperature can act as a proxy for stable boundary layer and low vertical mixing (Barmpadimos et al., 2011).

3.4 Performance of the ANN model

The ANN models were trained to predict the upper limit of the hourly road contributions to PM_{10} concentrations, and after training, they were tested for the prediction on the test data. The test performance of the models was evaluated, and the results show that the models performed very well on the test data. Table 4 shows the statistical performance of the models for each site. The predicted PM_{10} concentrations show good correlation with their corresponding observations.

Table 4 Performance statistics of the ANN predictions.

Site	FAC2	NMB	RMSE	R	COE	IOA
BT4	0.83	0.02	11.40	0.75	0.39	0.70
CD3	0.73	0.07	11.39	0.63	0.22	0.61
CR4	0.57	−0.01	12.51	0.60	0.28	0.64

[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
GR5	0.40	0.11	6.97	0.77	0.29	0.65
GR8	0.83	0.01	21.73	0.68	0.34	0.67
HK6	0.57	−0.02	9.09	0.70	0.28	0.64
IS2	0.73	0.01	9.97	0.86	0.50	0.75
KC2	0.94	−0.01	7.70	0.78	0.41	0.71
KC5	0.80	0.02	7.57	0.61	0.19	0.59
MY1	0.95	−0.01	11.79	0.76	0.38	0.69

The ANN models performed well on the test data with 57~~%~~–95% of their predictions falling within the factor of two of the estimated upper limit of road contributions of PM_{10} as shown by the FAC2 values in Table 4. Only at GR5 that about 40% of the predictions are within the factor of two of the road PM_{10} contributions. This low performance might be attributed to the averaging period used in collecting the data at this site which is twenty-four-hour average instead of hourly averages as it was obtained at the remaining sites. The daily averages are repeated for the 24 h of the day to get the hourly values. Therefore, the models tried to estimate the hourly variations within the predictors that did not exist in the response variable hence the low performance. The FAC2 values are well above the minimum of 50% recommended to the DEFRA UK (Derwent et al., 2010) for the acceptance of an air quality model. The models showed low bias in their predictions. However, there is slight under-prediction of the road contributions at some of the sites, as indicated by the negative sign of the NMB values.

Derwent et al. (2010) recommended that for an air quality model to be accepted it should satisfy the minimum requirement of NMB values in the range between −0.2 and +0.2. The NMB values obtained here for all the sites are nearly zero which shows that they performed way above the minimum requirement. Also as in the case of FAC2 values, the NMB value for the GR5 site is slightly higher than the corresponding values at the other sites. The coefficient of correlation between them is between 0.61 and 0.86 showing good agreement between the model predictions and the observations of the road contributions. The models can estimate the road contributions with about 20–50% more accuracy than the mean of the road contributions as indicated by the Coefficient of Efficiency (COE) values. The RMSE values range between 6.97 and~~–~~ 21.73 μg/m^3 depending on the hourly average PM_{10} concentrations at the sites. However, a perfect model should have 0 RMSE value, while the higher the RMSE values shown by the model, the less accurate its predictions will be. Although the RMSE is sensitive to extreme values, it reveals the actual size of the error produced by the model (Willmott, 1982, 1981).

Also, the index of the agreement shows that the prediction of the models is in good agreement with the observed data with values ranging between 60 and 75%. The relationships between the predicted and observed values were further analysed using time variation plots as to ascertain how well the models captured the temporal variations in the observations as shown in Fig. 4. The hourly variation plots demonstrate that the models perform extremely well in capturing the hourly variation of the PM_{10} observations at the sites and all the predictions are within the 95% confidence intervals of the observation as shown in Fig. 4.

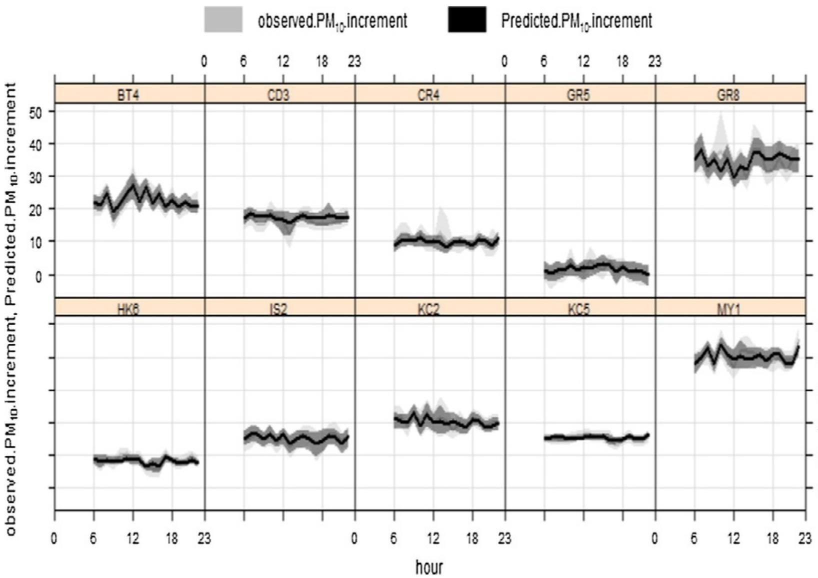

Fig. 4 Hourly time variation plots comparing the predicted and observed upper limit of road traffic contribution to the PM_{10} concentrations at 10 sites.

The predictions of the average daily traffic contributions also followed the observations closely and were all within the 95% confidence intervals. Although there is no much difference in the daily contributions, there are days with higher and lower contributions at some sites. For example at BT4 and GR8 sites, Thursdays show higher and lower contributions respectively. At MY1 site, Wednesdays and Saturdays show lower contributions than the remaining days. At CD3 the models overestimated the weekend contributions while they underestimated Friday contributions at CR4 site (see Fig. 5).

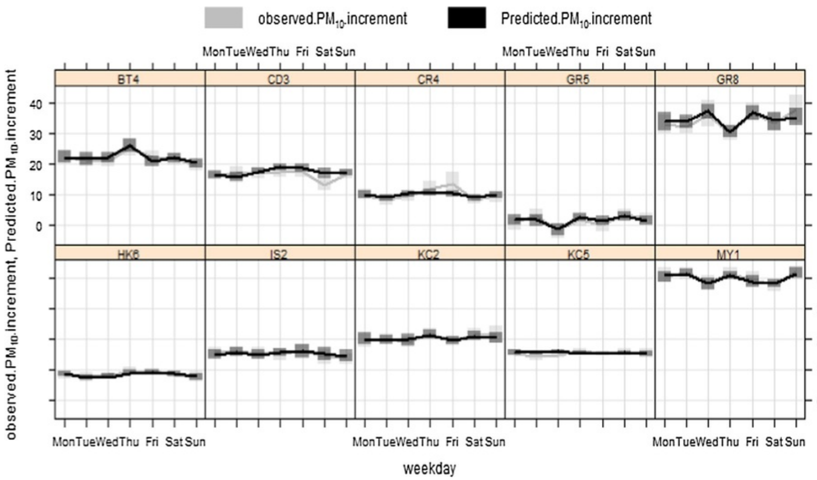

Fig. 5 Daily time variation plots comparing the predicted and observed upper limit of road traffic contribution to the PM_{10} concentrations at 10 sites.

The scatter plots in Fig. 6 show the graphical correlation between the observed and predicted road contributions. It could be seen that the models over predicted the lower contributions at most of the sites. However they predicted the medium to higher contributions more accurately. Also, there is slight underestimation at HK6 and CR4. The models performed better at BT4, IS2, KC2, and MY1 where most of the predictions at these sites fall within the FAC2 boundaries demarcated by the two dashed lines in each of the plots in Fig. 6.

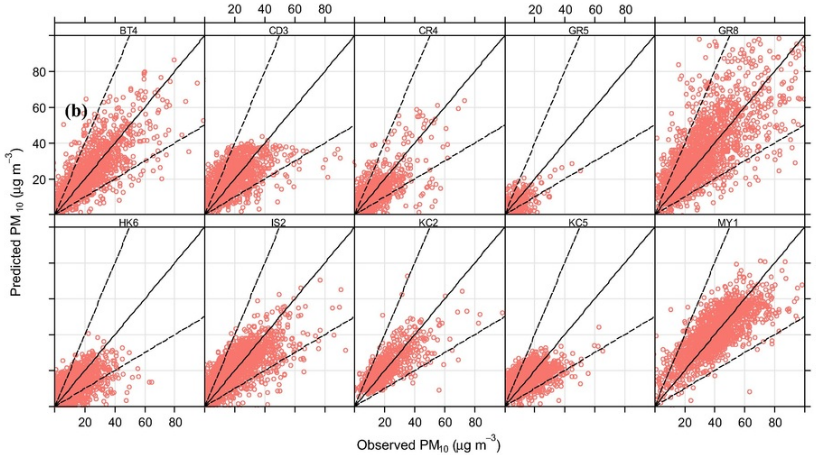

Fig. 6 Scatter plots comparing the predicted and observed upper limit of road traffic contribution to the PM_{10} concentrations at 10 sites.

The boxplots in Fig. 7 show that the ANN models predicted the statistical values of the contributions i.e. minimum, mean and maximum values, the lower and upper quantiles of the observed road traffic contributions accurately. Although these statistics are conservative, they give information on how the models captured the distribution of the observed contribution of road traffic to the PM_{10} concentrations (see Fig. 8 (This should be Fig. 7 or be removed if the first reference is sufficient)).

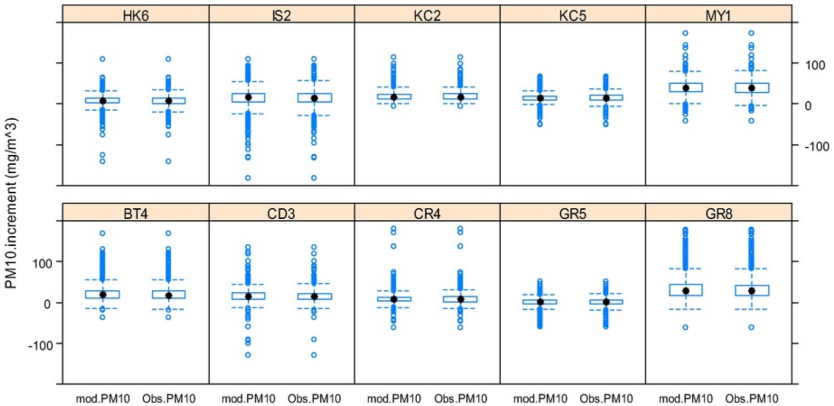

Fig. 7 Box and whisker plot comparing the predicted (mod.PM_{10}) and observed (obs.PM_{10}) upper limit of road traffic contribution to the PM_{10} concentrations at 10 sites.

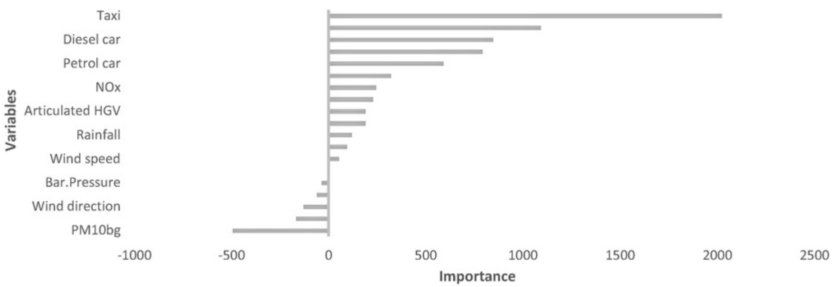

Fig. 8 Variable importance in the ANN models.

The contribution of the input variables to the output of the models was estimated using the connection weight approach (Olden et al., 2004). The results obtained for the MY1 site is shown in Fig. 6 (This should be Fig. 8). It could be seen that the most contributing variables are the hourly PM_{10} emission rates of the vehicles among which the London taxi, diesel car, LGV and the HGVs contributed the most. This relationship is expected since these vehicles are known to have higher particle emissions. (Fig 8 should be inserted here if possible) The results agree with the PCA analysis carried out during the preprocessing of the model inputs where the PCs with higher variability in the data set correlated very well with the traffic variables. The second most important variables are the background PM_{10} and oxides of nitrogen and lastly the meteorological variables. The background PM_{10} should have direct relationships with the roadside PM_{10} since the later is just an addition over the former due to the engine emissions, wear and tear of the vehicle parts, brakes and road dust. The meteorological variables play a vital role in the dispersion of the PM_{10} within the urban environment. Stability of boundary layer, street canyon recirculation, resuspension of particles, deposition, and chemical transformations of the emitted particles are all controlled by the meteorological variables such as temperature, wind speeds and directions, rainfall and relative humidity to mention a few. Therefore, the result of the variable contribution in the ANN model development boosted our confidence on its estimates since it describes the underlaying relationships between the variable involved in the prediction. Also, this result could help in identifying the vehicles that emit higher particles for the purpose of emission managements and control and in turn the ANN models could be considered as the air quality management tools. Although the result of this study in using ANN methods to predict road traffic contribution to particulate matter is promising, its accuracy will largely depend on the assumptions on the period of traffic activity, wind speed threshold at which the PM_{10} concentration is considered to be dominated by traffic sources and the wind sectors related to the roads.

4 Conclusions

The method for estimating and predicting the upper limit of road traffic contribution to the urban particulate matter has been demonstrated and shows that the use of twin studies in conjunction with the bivariate polar plots could lead to more accurate estimates of the contribution of road traffic to the roadside concentrations of particulate matter. The influence of local and far distance emission sources on the concentration was also highlighted. The results showed that using twin studies without filtering the data based on the polar plot analysis underestimates the road contribution by almost 50%. It was also found out the there was not much variation in the magnitudes of the contribution of road traffic and other sources. However the frequency of observations of low winds and wind direction related to the roads is much higher than the frequency of observations under higher winds and other wind directions. For example at lower winds the frequency of observations was found to be between 27% ~~to~~and 44% of the total observation at the sites, while the frequency of the PM_{10} observation due to other sources constitutes only between 2% ~~to~~and 4%. The percentage mean upper limit contribution of the road traffic to the hourly average roadside concentrations was estimated to be between 24% ~~to~~and 64%. The ANN models performed extremely well in predicting the hourly road traffic contributions, and the predictions agreed well with the observations. The analysis of variable importance confirmed the earlier results obtained by the use of principal component analysis. It shows that the traffic emission variables were found to have contributed immensely to the accuracy of the models. The results of analysis of variable importance demonstrates that the traffic emission variables are the most influential variables in the prediction, therefore it provides an opportunity for the models to be used as air quality management tools where the sensitivity of the models to the changes in the emission could be used to test the effectiveness of an air quality management scenario related to traffic technology, composition or volume.

Acknowledgements

The authors would like to thank Petroleum Technology Development Fund (PTDF) for funding this research and also the London Air Quality network and Environmental research group at Kings College London for their help in providing data and Openair package for the data analysis.

References

Abdel-Aty M., Ekram A.-A., Huang H. and Choi K., A study on crashes related to visibility obstruction due to fog and smoke, *Accid. Anal. Prev.* **43**, 2011, 1730–1737.

Anderson J., Thundiyil J. and Stolbach A., Clearing the air: a review of the effects of particulate matter air pollution on human health, *J. Med. Toxicol.* **8**, 2012, 166–175.

Aurelie C. and Harrison R.M., Comparison between SMPS, nano-SMPS and Epiphaniometer Data at an Urban Background Site (Bloomsbury) and a Roadside Site (Marylebone Road), 2005.

Barmpadimos I., Hueglin C., Keller J., Henne S. and Prévôt A., Influence of meteorology on PM 10 trends and variability in Switzerland from 1991 to 2008, *Atmos. Chem. Phys.* **11**, 2011, 1813–1835.

Bishop C.M., Neural Networks for Pattern Recognition, 1995, Oxford University Press, Inc..

Brunekreef B., Beelen R., Hoek G., Schouten L., Bausch-Goldbohm S., Fischer P., Armstrong B., Hughes E., Jerrett M. and van den Brandt P., Effects of long-term exposure to traffic-related air pollution on respiratory and cardiovascular mortality in the Netherlands: the NLCS-AIR study, *Res. Rep. Health Eff. Inst.* **139**, 2009, 5–71.

Carslaw D.C. and Beevers S.D., Characterising and understanding emission sources using bivariate polar plots and k-means clustering, *Environ. Modell. Softw.* **40**, 2013, 325–329.

Carslaw D.C., Beevers S.D., Ropkins K. and Bell M.C., Detecting and quantifying aircraft and other on-airport contributions to ambient nitrogen oxides in the vicinity of a large international airport, *Atmos. Environ.* **40**, 2006, 5424–5434.

Chiang H.-L., Lai Y.-M. and Chang S.-Y., Pollutant constituents of exhaust emitted from light-duty diesel vehicles, *Atmos. Environ.* **47**, 2012, 399–406.

de Gennaro G., Trizio L., di Gilio A., Pey J., Perez N., Cusack M., Alastuey A. and Querol X., Neural network model for the prediction of PM10 daily concentrations in two sites in the Western Mediterranean, *Sci. Total Environ.* **463–464**, 2013, 875–883.

Derwent D., Fraser A., Abbott J., Jenkin M., Willis P. and Murrells T., Evaluating the performance of air quality models, In: (DEFRA) D.F.E.F.A.R.A., (Ed), *Department for Environment Food and Rural Affairs (DEFRA): Department for Environment Food and Rural Affairs (DEFRA),* 2010.

Ding S., Li H., Su C., Yu J. and Jin F., Evolutionary artificial neural networks: a review, *Artif. Intell. Rev.* **39**, 2011a, 251–260.

Ding S., Su C. and Yu J., An optimizing BP neural network algorithm based on genetic algorithm, *Artif. Intell. Rev.* **36**, 2011b, 153–162.

Elangasinghe M.A., Singhal N., Dirks K.N. and Salmond J.A., Development of an ANN-based air pollution forecasting system with explicit knowledge through sensitivity analysis, *Atmos. Pollut. Res.* **5**, 2014a.

Elangasinghe M.A., Singhal N., Dirks K.N., Salmond J.A. and Samarasinghe S., Complex time series analysis of PM10 and PM2.5 for a coastal site using artificial neural network modelling and k-means clustering, *Atmos Environ.* **94**, 2014b, 106–116.

Gouriou F., Morin J.P. and Weill M.E., On-road measurements of particle number concentrations and size distributions in urban and tunnel environments, *Atmos. Environ.* **38**, 2004, 2831–2840.

Harrison R.M., Jones A.M. and Barrowcliffe R., Field study of the influence of meteorological factors and traffic volumes upon suspended particle mass at urban roadside sites of differing geometries, *Atmos. Environ.* **38**, 2004, 6361–6369.

He H.-D., Lu W.-Z. and Xue Y., Prediction of particulate matter at street level using artificial neural networks coupling with chaotic particle swarm optimization algorithm, *Build. Environ.* **78**, 2014, 111–117.

Jones A. and Harrison R., Estimation of the emission factors of particle number and mass fractions from traffic at a site where mean vehicle speeds vary over short distances, *Atmos. Environ.* **40**, 2006, 7125–7137.

Kassomenos P., Vardoulakis S., Chaloulakou A., Grivas G., Borge R. and Lumbreras J., Levels, sources and seasonality of coarse particles (PM 10–PM 2.5) in three European capitals – implications for particulate pollution control, *Atmos. Environ.* **54**, 2012, 337–347.

Ketzel M., Wåhlin P., Berkowicz R. and Palmgren F., Particle and trace gas emission factors under urban driving conditions in Copenhagen based on street and roof-level observations, *Atmos. Environ.* **37**, 2003, 2735–2749.

KUHN, M. 2012. The Caret Package.

Lawal A.O., Zhang M., Dittmar M., Lulla A. and Araujo J.A., Heme oxygenase-1 protects endothelial cells from the toxicity of air pollutant chemicals, *Toxicol. Appl. Pharmacol.* **284**, 2015, 281–291.

Lima A.R., Cannon A.J. and Hsieh W.W., Nonlinear regression in environmental sciences by support vector machines combined with evolutionary strategy, *Comput. Geosci.* **50**, 2013, 136–144.

LondonAir, London Air Quality Network [Online], Available: 2013, (accessed 03/04/2013 2013).

Masiol M. and Harrison R.M., Quantification of air quality impacts of London Heathrow Airport (UK) from 2005 to 2012, *Atmos. Environ.* **116**, 2015, 308–319.

Midas land surface M.O., Met Office Integrated Data Archive System (MIDAS) Land and Marine Surface Stations Data (1853-current) [Online], Available: 2013, (accessed 24/06/2013 2013).

Olden J.D. and Jackson D.A., Illuminating the “black box”: a randomization approach for understanding variable contributions in artificial neural networks, *Ecol. Model.* **154**, 2002, 135–150.

Olden J.D., Joy M.K. and Death R.G., An accurate comparison of methods for quantifying variable importance in artificial neural networks using simulated data, *Ecol. Model.* **178**, 2004, 389–397.

Pant P. and Harrison R.M., Estimation of the contribution of road traffic emissions to particulate matter concentrations from field measurements: a review, *Atmos. Environ.* **77**, 2013, 78–97.

Pokorná P., Hovorka J., Klán M. and Hopke P.K., Source apportionment of size resolved particulate matter at a European air pollution hot spot, *Sci. Total Environ.* **502**, 2015, 172–183.

R Development Core Team, 2015. R 3.2. 1. R Project for Statistical Computing Vienna, Austria.

Ragosta M., D’Emilio M. and Giorgio G.A., Input strategy analysis for an air quality data modelling procedure at a local scale based on neural network, *Environ. Monit. Assess.* **187**, 2015, 307.

Russo A., Raischel F. and Lind P.G., Air quality prediction using optimal neural networks with stochastic variables, *Atmos. Environ.* **79**, 2013, 822–830.

Singh K.P., Gupta S. and Rai P., Identifying pollution sources and predicting urban air quality using ensemble learning methods, *Atmos. Environ.* **80**, 2013, 426–437.

Taspinar F., Improving artificial neural network model predictions of daily average PM10 concentrations by applying principle component analysis and implementing seasonal models, *J. Air Waste Manage. Assoc.* **65**, 2015, 800–809.

TFL, Transport emissions roadmap (TERM), In: London M.O., (Ed), *Transport for London,* 2014.

Tomlin A.S., Smalley R.J., Tate J.E., Barlow J.F., Belcher S.E., Arnold S.J., Dobre A. and Robins A., A field study of factors influencing the concentrations of a traffic-related pollutant in the vicinity of a complex urban junction, *Atmos. Environ.* **43**, 2009, 5027–5037.

UK-AIR, Department fro Environment Food and Rural Affairs Data Archive [Online], Available: 2013, (accessed 03/04 2013).

Uria-Tellaetxe I. and Carslaw D.C., Conditional bivariate probability function for source identification, *Environ. Modell. Softw.* **59**, 2014, 1–9.

Willmott C.J., On the validation of models, *Phys. Geogr.* **2**, 1981, 184–194.

Willmott C.J., Some comments on the evaluation of models, *Bull. Am. Meteorol. Soc.* 1982.

Yang L., Zhou X., Wang Z., Zhou Y., Cheng S., Xu P., Gao X., Nie W., Wang X. and Wang W., Airborne fine particulate pollution in Jinan, China: concentrations, chemical compositions and influence on visibility impairment, *Atmos. Environ.* **55**, 2012, 506–514.

Highlights

- The dispersion of roadside PM_{10} concentrations was investigated.
- The contribution of road traffic on major roads to PM_{10} levels are estimate.
- The roads contributes 24–62% of the average roadside PM_{10} concentrations.
- The ANN models performed well in predicting the road contributions to PM_{10}.
- Traffic variables were the most contributing variables to the output of the models.

Queries and Answers

Query: Your article is registered as a regular item and is being processed for inclusion in a regular issue of the journal. If this is NOT correct and your article belongs to a Special Issue/Collection please contact h.beulah@elsevier.com immediately prior to returning your corrections.
Answer: The paper was submitted to a regular issue

Query: The author names have been tagged as given names and surnames (surnames are highlighted in teal color). Please confirm if they have been identified correctly.
Answer: The surnames have been identified correctly

Query: Please check the edits made in the sentence 'The circles indicate...', and correct if necessary.

Answer: Yes

Query: Reference 'DEFRA (2013)' is cited in the text but not provided in the reference list. Please provide it in the reference list or delete this citation from the text.

Answer: DEFRA 2013. Emissions of Air Quality Pollutants 1970 - 2011 *In:* DEPARTMENT FOR ENVIRONMENT, F. A. R. A. U. (ed.). AQPI Summary Report.

Query: To maintain sequential order, tables have been renumbered. Hence, their corresponding citations have also changed throughout the text. Please check, and correct if necessary.

Answer: Yes

Query: Please note that 'Figs. 5 and 8' were not cited in the text. Please check that the citations suggested by the copyeditor are in the appropriate place, and correct if necessary.

Answer: Figure 5 is okay but Figure 8 is not correct

Query: Please provide a volume number, or issue number, or page range for reference 'Willmott (1982)'.

Answer: Willmott, C.J., 1982. Some comments on the evaluation of model performance. *Bulletin of the American Meteorological Society*,*63*(11), pp.1309-1313.